

Overeenkomsten tussen DIP en DAP

Toine Heuvelmans
Audio Design
Fac. KMT, Hogeschool voor de Kunsten Utrecht
toine@student-kmt.hku.nl

Als beeldeffecten op geluid worden toegepast, of vice versa, wat zal dan het resultaat zijn? Is het eigenlijk wel mogelijk om dit te doen? Alleen aan door het vinden van overeenkomsten zou dit kunnen lukken, maar dan moeten die er wel zijn...

DSP richt zich op analyse en bewerking van digitale signalen. Aangezien de belangrijkste menselijke zintuigen zicht en gehoor zijn, is veel DSP gericht op beeld (*DIP*, zie kader 1) en geluid (*DAP*).

De meest bekende vorm waarin wij *bewust* te maken krijgen met deze twee technieken is plugins. Dit zijn kleine gespecialiseerde extensies voor grote programma's. Denk aan de filters in programma's als Photoshop of GIMP, en de effecten in programma's als Logic of GarageBand. Er zijn in de afgelopen jaren voor beide gebieden honderden plugins ontwikkeld, door zowel grote bedrijven als amateurs. Hoewel ze tegelijk met de PC steeds geavanceerder en sneller worden, blijven de ideeën erachter vrijwel hetzelfde. Inspiratie voor iets nieuws dient wellicht ergens anders gezocht te worden dan binnen hetzelfde gebied.

Er zijn een aantal parameters van signalen beschreven die algemeen zijn binnen heel DSP. Aangezien DIP en DAP de meest belangrijke gebieden zijn, vallen hiertussen via deze parameters misschien vrij eenvoudig overeenkomsten te vinden.

Digital Audio Processing

Overzicht

Geluidspecifieke DSP heet Digital Audio Processing. DAP kennen we voornamelijk als de geluidseffecten in muziekprogramma's, zoals filters, EQ, galm etc. Dit zijn echter alleen nog maar bewerkingen. Een van de belangrijkste onderdelen is namelijk analyse. Het gedrag van geluid wordt geanalyseerd, om bijvoorbeeld erachter te komen wat we wel horen en wat niet. Dit is terug te vinden in alle meters en andere weergaven die we in muziekprogramma's zien, maar ook bijvoorbeeld in de akoestiek en psycho-akoestiek. Verder gaat DAP over onderwerpen als compressie (van zowel dynamiek als de data zelf), opslag en overdracht van audio.

kader 1. afkortingen

DSP:	Digital Signal Processing <i>Het analyseren en bewerken van digitale signalen</i>
DIP:	Digital Image Processing <i>Het analyseren en bewerken van digitaal beeld</i>
DAP:	Digital Audio Processing <i>Het analyseren en bewerken van digitaal geluid</i>

Representatie

Audio wordt digitaal gerepresenteerd als een ééndimensionale stroom amplitudes in de tijd, ook wel *samples* genoemd (zie fig. 1).

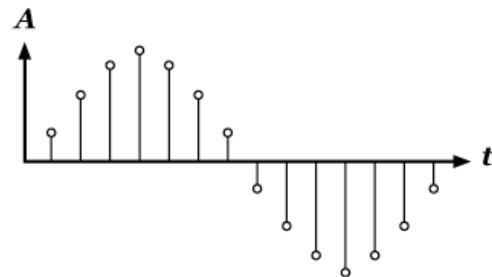


fig. 1

De hoeveelheid samples per seconde wordt aangegeven door de samplerate. Vrijwel altijd ligt deze samplerate in de duizenden (meestal 44100). Aan zoveel amplitudes per seconde kunnen we dus vrij weinig aflezen, eigenlijk alleen een globale amplitude in een bepaald deel van de stroom. Toch horen we van alles en nog wat door elkaar, het een hard, het ander zacht. Dit komt omdat we luisteren naar een optelling van een hele hoop golfvormen, elk met een andere amplitude en fase. Een sample geeft dus de amplitudes van meerdere golfvormen op één zelfde tijdstip weer.

Ons oor ontvangt op deze manier golfvormen, aangezien het trommelvlies maar in één positie kan staan op een bepaald tijdstip. Daarna wordt het geluid door onze hersenen geanalyseerd zodat wij meerdere dingen tegelijk kunnen horen. Ons oor vangt dus ongeanalyseerd geluid op. Hier kunnen we uit opmaken dat digitale audio ook ongeanalyseerd is. Dit verklaart ook waarom analyse binnen DAP een heel belangrijke rol speelt.

Digital Image Processing

Alle DSP die te maken heeft met beeld, valt onder Digital Image Processing. DIP heeft zich wat langzamer kunnen ontwikkelen dan DAP, omdat opgeslagen beeld veel informatie bevat, en dus

veel ruimte kost. Nu pas beginnen PC's genoeg ruimte en snelheid te krijgen om er echt mee aan de slag te kunnen.

Overzicht

DIP is breder dan men op het eerste gezicht zou denken. Beeld is er namelijk in veel verschillende vormen, soms zelfs onzichtbaar voor ons. Denk bijvoorbeeld aan het maken van een Röntgenfoto, of wat infrarood camera's weergeven. Veel technieken zijn, net zoals deze, gericht op het weergeven of versterken van iets dat we in eerste instantie niet of nauwelijks zien.

Commerciële DIP vinden we allereerst in compressietechnieken. Om het ruimtegebrek van PC's te compenseren, kan een hoop overbodige informatie uit beeld weg worden gefilterd. Niet alleen op PC's wordt dit toegepast, ook bij TV, film, games, video-telefonie etc.

En ook zijn er zeer uiteenlopende plugins, waarvan veel zich richten op het versterken van beeld, zoals contrast- en contourversterking, en opschoning zoals ruis- en blurvermindering.

Representatie

Digitaal beeld is wat betreft digitale signalen een vreemde eend in de bijt. Beeld is namelijk een meting van parameters in ruimte, in tegenstelling tot andere signalen, die een meting zijn van parameters in tijd.

Beeld is namelijk een fragmentopname, of beter gezegd een fragmentanalyse. De digitale representatie komt overeen met wat onze hersenen op een bepaald tijdstip maken van inkomende lichtgolven (afgezien van het feit dat we maar op een klein punt kunnen focussen).

Digitaal beeld is dus al geanalyseerd. Hierdoor bevat het veel informatie, wat er voor zorgt dat het ook aanzienlijk meer ruimte kan kosten om het op te slaan.

De analyse is een tweedimensionaal vlak waar in een rooster kleuren staan, ook wel *pixels* genoemd. De combinatie van de plaats in het rooster en de kleur van de pixel geeft uiteindelijk het beeld weer.

Representatie en benadering

Signalen bestaan allemaal uit golfvormen. Er zijn een aantal belangrijke parameters die we goed moeten begrijpen, willen we bevatten wat digitale signalen precies weergeven, zodat we ze op de juiste manier kunnen benaderen.

Intensiteit

De intensiteit van een golfvorm wordt aangegeven door de amplitude. Hoe hoger deze amplitude is, hoe intenser we de golfvorm ervaren.

Bij geluid wordt deze intensiteit gemeten door ons trommelvlies. Digitaal geluid wordt precies weergegeven als de uitwijkingen van ons trommelvlies. Hoe hoger de amplitude van een sample is, hoe verder hij uitwijkt.

Bij digitaal beeld kunnen de pixels gezien worden als verschillende kleuren met een bepaalde intensiteit, op een bepaalde plaats in een vlak.

Kleur wordt in een pixel aangeduid als een combinatie van kleuren, vaak rood, groen en blauw, net zoals bij de kegeltjes in ons oog. De intensiteit van deze kleuren bepaalt welke kleur we zien. Deze intensiteiten zijn om te rekenen naar een representatie die beter weergeeft hoe een kleur in elkaar zit. Door middel van *Hue* (tint), *Saturation* (intensiteit van tint), en *Brightness* (intensiteit van belichting) wordt de manier waarop een kleur wordt gevormd beter benaderd.

De kleur van object wordt bepaald door de mate waarin frequenties erdoor worden gereflecteerd. Als een object bijvoorbeeld alleen frequenties van rood licht reflecteert, zien we het object als rood. Dit is de tint (*Hue*) van het object.

Als de andere kleuren gelijkmatig ook een beetje gereflecteerd worden, wordt er in totaal dus meer licht gereflecteerd, en zal de kleur van het object meer naar wit (de aanwezigheid van alle kleuren) neigen. Rood contrasteert dan minder met de andere kleuren, en zal dus minder fel zijn. De intensiteit van de tint (*Saturation*) neemt af.

Als het object minder wordt belicht (*Brightness*), zal er ook minder licht gereflecteerd worden, en zal de kleur meer naar zwart (het ontbreken van alle kleuren) neigen. Het hangt dus uiteindelijk van de belichting af hoe intens we de kleur ervaren, oftewel wat de amplitude is.

Frequentie

Met frequentie wordt aangeduid hoe vaak per tijdseenheid een golfbeweging wordt gemeten. Een hoorbare frequentie ervaren we als een toon, een zichtbare frequentie als een kleur.

Een aantal zichtbare frequenties bij elkaar zorgt ervoor dat we een *object* kunnen zien. De frequenties hoeven niet te veranderen om het object te kunnen onderscheiden. Als er nog meer frequenties bij komen, van de omgeving waarin het object staat, dan zien we nog steeds het object. Denk maar aan een foto, waarin ook geen enkele frequentie verandert.

Als we nu een aantal hoorbare frequenties hebben die niet in de tijd veranderen, horen we nog steeds een toon, alleen een wat complexere. Neem bijvoorbeeld een orkest, waarvan ieder instrument continu dezelfde frequenties uitstuurt, net als bij een foto. Een enkel instrument zal er niet in te onderscheiden zijn.

Pas als de frequenties van het orkest veranderen in de tijd, kunnen we instrumenten onderscheiden. Dit komt omdat onze hersenen

dan *timbre* onderscheiden. *Timbre* is de verhouding (en het verloop) van amplitudes van de partialen in een *klank*. In de klank van een instrument hangen de frequenties met hun amplitudes op een bepaalde manier samen, waar ons gehoor zich op kan focussen. Als zo'n samenhang wordt onderscheiden, kunnen wij die specifieke klank ook onderscheiden. Wat ook een grote rol speelt bij de herkenning van een instrument, is de eerste milliseconden van de klank (de *attack*), waarin veel informatie zit die kenmerkend is voor dat instrument. Zonder *attack* zouden we niet eens het verschil horen tussen een viool en een trompet.

Een momentopname is eigenlijk een frequentieanalyse. Alle frequenties op dat moment staan erin aangegeven. Het verschil tussen beeld en geluid, de reden waarom we in een momentopname een object kunnen onderscheiden en een klank niet, zit hem in de andere parameters in de analyse.

Zowel bij beeld- als geluidanalyse is ook de intensiteit van de frequenties "uit te lezen". Bij geluid is er nog één extra parameter aanwezig, namelijk fase. Dit omschrijft de plaats in de golfbeweging van een frequentie. Hier merken we echter niet echt veel van.

Bij beeld is er ook één extra parameter (of eigenlijk twee), namelijk plaatsing van de pixel (in x- en y-coördinaten). Hier merken we wel degelijk wat van. Frequenties worden door deze parameter geordend in een vlak, waardoor het eenvoudiger voor ons wordt om iets erin te herkennen. Als we één van de twee coördinaten weg zouden halen, bijvoorbeeld het de x-as, dan zien we één lijn van pixels. Dit is even onherkenbaar als een momentopname van geluid (waar de parameter *tijd* eigenlijk uit is verwijderd). Als we dus een goede analyse van een klank willen maken, dan is het dus heel belangrijk dat de parameter tijd ook aanwezig is.

Tijd en causaliteit

Het blijkt dus dat voor beeld de parameter *plaatsing* heel belangrijk is, en voor geluid de parameter *tijd*. Dit is ook terug te vinden in de manier waarop het digitale signaal benaderd wordt. Elke *sample* staat op een ander punt in de tijd, elke *pixel* staat op een andere plaats in het vlak. Aangezien een vlak tweedimensionaal is, en tijd slechts ééndimensionaal, heeft een pixel meer directe "buren" dan een sample (zie fig. 2).

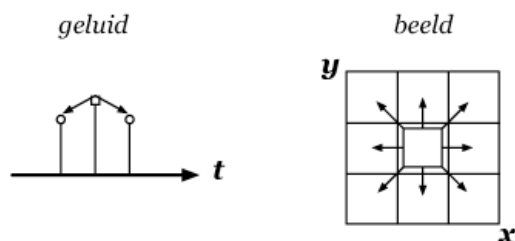


fig. 2

Het verband tussen een pixel of sample en zijn buren is belangrijk voor het karakter van een signaal. Als de waarde van een sample veel verschilt met die van zijn buren, houdt dit in dat er een snelle golfbeweging wordt beschreven, oftewel dat de frequentie hoog ligt. Als een pixel veel verschilt met zijn buren, contrasteert het met zijn omgeving, en zal het dus duidelijk zichtbaar zijn. Aangezien een pixel rondom acht buren heeft, kan dit contrast op verschillende manieren worden geïnterpreteerd – als een punt, lijn of rand.

Contrasterende frequenties in geluid kunnen niet zulke geometrische indrukken geven, maar het kan wel uitmaken hoe goed een frequentie is te horen. Hierbij speelt tijd echter weer een grote rol. Als een frequentie een zelfde soort amplitudeverloop heeft als andere frequenties, zal hij snel worden gehoord als een partiaal, een boventoon die deel uitmaakt van een klank. Als de frequentie echter een heel ander amplitudeverloop heeft, zal hij los van de klank staan, en te onderscheiden zijn.

Als de waarden van pixels in de tijd veranderen, zoals bij video, is het mogelijk dat het contrast zich tussen andere pixels vormt. Dit lijkt dan een verplaatsing van het contrast, en zal resulteren in het zien van beweging.

Analyses

Een analyse vertelt iets over een signaal op een bepaald moment (de eerder genoemde momentopname). Als meerdere analyses achter elkaar worden gezet, kan een verloop in de tijd worden weergegeven. Bij het analyseren van klank is dit van groot belang, aangezien amplitudes op een bepaalde manier in de tijd verlopen.

Er kunnen verschillende soorten analyses van beeld en geluid worden gemaakt, elk specifiek gericht op het uitlichten van één of meerdere parameters. Sommige technieken kunnen op zowel beeld als geluid toegepast worden. Maar wat analyseren ze dan, aangezien de signalen zo verschillend zijn?

Histogram

Een histogram is een grafische weergave van een bepaalde verdeling. In beeldbewerking wordt vaak een intensiteitshistogram gebruikt om aan te geven hoeveel van een bepaalde intensiteit (globaal of specifiek van rood, groen of blauw) aanwezig is (zie fig. 3).

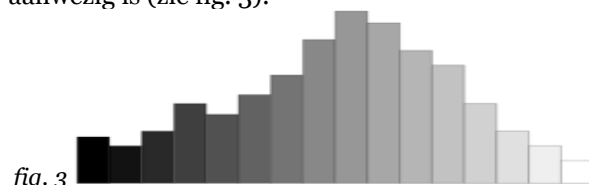
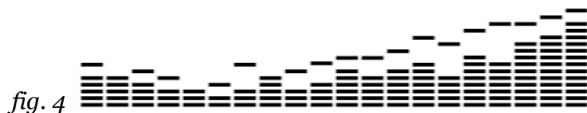


fig. 3

Hier is dus globaal uit af te lezen hoe “intens” een afbeelding is. Om een intensiteit histogram te maken hoeft alleen maar per pixel gekeken te worden wat de intensiteit is. Een balk in het histogram geeft de hoeveelheid pixels met die intensiteit (of intensiteit binnen een bepaald bereik) weer. Als een afbeelding helemaal zwart zou zijn, zou alleen de eerste balk zichtbaar zijn, en de waarde hiervan zou gelijk zijn aan het volledige aantal pixels in de afbeelding. Zo'n histogram zien we ook vaak in muziekprogramma's, zoals iTunes of WinAmp.



Deze geven alleen niet het aantal frequenties per bepaalde intensiteit, maar andersom, de intensiteit per frequentie (of frequentieband, zoals in fig. 4). Het is stukken ingewikkelder om een dergelijk histogram (of *spectrumanalyse*) te krijgen. Het geluid moet namelijk geanalyseerd worden, of zoals de naam omschrijft, het spectrum moet geanalyseerd worden. Analoge apparaten gebruiken hiervoor filters die alleen een bepaalde frequentieband doorlaten, en meten de intensiteit van de output van het filter.

Digitaal wordt gebruik gemaakt van een van de belangrijkste technieken binnen heel DSP, de Fourier Analyse.

Fourier Analyse

Digitaal is veel meer mogelijk, zo ook uitgebreidere wiskunde. De Fourier Analyse maakt uitbundig gebruik van wiskunde, en kan dus alleen digitaal geautomatiseerd worden. Volgens Fourier bestaat elk willekeurig signaal uit een grondtoon met een oneindig aantal *harmonischen* (frequenties in een bepaalde verhouding), met elk een eigen amplitude en fase. Met behulp van de Fourier Transformatie kunnen deze harmonischen met hun parameters worden geëxtraheerd, en kan het in een histogram worden weergegeven. De transformatie is omkeerbaar, waardoor het mogelijk wordt de frequenties zeer effectief te manipuleren.

Er is ook een Fourier Transformatie voor beeld. Het benodigde signaal voor deze transformatie is niet de afbeelding zelf, maar de intensiteiten van de pixels in elke lijn. Per lijn wordt een analyse gemaakt, waarin snelle contrastveranderingen uit worden gedrukt als hoge frequenties, en grote contrastverschillen als hoge amplitudes. De intensiteiten van het histogram dat hier uiteindelijk uit komt, worden vervolgens weer als pixels weergegeven, op dezelfde plaats als de geanalyseerde lijn. Zodoende ontstaat een combinatie van histogrammen, met dezelfde afmetingen als de afbeelding zelf.

Bewerkingen

Sommige signaalbewerkingen komen zowel binnen DIP als DAP voor, andere maar in één van de twee. Het hangt af van de benodigde signaalparameters op welk soort signalen een bepaalde bewerking kan worden toegepast.

Convolutie

Convolutie is wel de meest belangrijke techniek binnen heel DSP. Enorm veel systemen zijn met behulp van convolutie (na) te maken.

Een *stelsel* is een signaalbewerking met bepaalde parameters. Mits deze parameters niet veranderen, heeft het systeem bepaalde kenmerkende responsies. De frequentierespons geeft weer wat het systeem met de frequenties van het signaal doet. De staprespons geeft weer wat het systeem met snelle veranderingen binnen het signaal (hoge frequenties) doet. Als laatste is er de impulsrespons, die weergeeft wat het systeem met een enkele sample doet. Met behulp van deze responsies kan het systeem omschreven worden.

Convolutie neemt de impulsrespons van een systeem, en *convolueert* ("mengt") het met het invoersignaal. Dit geeft hetzelfde resultaat als wanneer het invoersignaal door het systeem wordt gehaald.

Wat betreft geluid is, naast standaard filters, zelfs een galm van een ruimte of een versterker een systeem. Het geeft een bepaald karakter aan het geluid. Door een enkele sample met de hoogste amplitude door zo'n systeem te sturen is die karakteristiek in een impulsrespons te vangen. Als vervolgens een signaal op de computer met die impulsrespons wordt geconvolueerd, zal het net klinken alsof het uit die ruimte of die versterker komt.

Een impulsrespons van geluid is een reeks factoren in de tijd. Elke sample wordt met iedere factor vermenigvuldigd, en op het tijdstip van die factor geplaatst. Het bepaalt eigenlijk hoeveel er nog te horen is van die ene sample op dat tijdstip.

Bij beeld wordt er niet gesproken van een impulsrespons, maar van een *kernel*. Een kernel bevat ook factoren, maar deze omschrijven de mate waarin naburige pixels bij elkaar moeten worden opgeteld. De som van deze fracties is uiteindelijk de uitvoer op de plaats van de centrale pixel. Het omschrijft dus eigenlijk hoeveel er op een plaats van zijn omgeving te zien is.

Filters

Het woord *filter* geeft aan dat een systeem bepaalde dingen doorlaat, en andere niet. Wat betreft DSP worden hiermee vrijwel altijd frequentiefilters bedoeld. De naamgeving van karakteristieke filters verschilt erg tussen beeld en geluid, omdat bij geluid de term *frequentie*

veel concreter is. In geluid beschrijft de naam van een filter dan ook vaak precies wat het filter met de frequenties doet. Beeldfilters worden meestal benoemd naar het effect dat het filter teweeg brengt. Een filter dat bijvoorbeeld lage frequenties doorlaat, maar hoge niet, heet in DAP een lowpass (laagdoorlaat) filter, en in DIP een blur filter. Het tegenovergestelde resulteert in respectievelijk een highpass (hoogdoorlaat) filter en een sharpen filter. Deze laatste versterkt eigenlijk de hoge frequenties ten opzichte van de lage, maar de karakteristiek is hetzelfde.

De beeldfrequenties die door deze filters worden behandeld, zijn de intensiteitsfrequenties die de Fourier Transformatie oplevert. Als naar frequentie in de zin van kleuren wordt gekeken, kan op een andere manier gefilterd worden. Bepaalde kleuren kunnen geheel weggelaten worden, met behulp van een *threshold* (grens). Vanaf die grens (er onder of erboven) valt alles weg.

Conclusie

Perceptueel is er een groot verschil tussen beeld en geluid. Beeld kan niet zonder ruimte, geluid kan niet zonder tijd. Onze ogen en oren vangen deze signalen dan ook op een verschillende manier op. De digitale representatie van beelden en geluidsignalen verschilt daarom ook aanzienlijk. Mar wat er daadwerkelijk gerepresenteerd wordt, verschilt niet. Signalen blijven frequenties met intensiteiten (amplitudes). Bij beeld zien we een analyse van deze frequenties, bij geluid een samenvoeging van de amplitudes. Met behulp van analysemethoden is het mogelijk om alle aanwezige parameters te extraheren. Zodoende kunnen bij zowel beeld als geluid dezelfde parameters bewerkt worden.

Er zijn veel bewerkingen die in beide gebieden voorkomen, hoewel dit op het eerste gezicht niet te merken is. Omdat de signalen perceptueel veel verschillen, lijken de resultaten van zo'n bewerking ook geheel niet op elkaar. Naamgeving van een bewerking gaat vaak aan de hand van het karakter van het resultaat. Pas als de daadwerkelijke handelingen van een bewerking bekeken worden, blijkt of het in beide gebieden toe te passen is. Bewerkingen op parameters frequentie en amplitude kunnen over het algemeen toegepast worden op beeld en geluid, maar bewerkingen op ruimte of tijd vrijwel niet (met uitzondering van tijdbewerkingen op video), omdat deze geen signaalparameters maar representatieparameters zijn. Frequentie en amplitude zijn dus de overeenkomsten tussen de signalen, de link tussen DIP en DAP.

Bronnen

Steven W. Smith: The Scientist and Engineer's Guide to Digital Signal Processing (1997). California Technical Publishing. www.dspguide.com

R. Fisher, S. Perkins, A. Walker, E. Wolfart: Hypermedia Image Processing Reference (2003). homepages.inf.ed.ac.uk/rbf/HIPR2

Wikipedia.org (WikiMedia Foundation) artikels:

Engels

- Digital Signal Processing
- Audio Signal Processing
- Digital Image Processing
- Numerical Analysis (category)
- Signal
- Color
- HSB

Nederlands

- Digitale beeldbewerking
- Timbre
- Klank

M. Green: Color in Trademark and Tradedress Disputes (2004). Visual Expert. www.visualexpert.com/Resources/color.html

J. M. Brayer: Experiments in Human Visual Perception
J. M. Brayer: Introduction to Fourier Transforms for Image Processing
www.cs.unm.edu/~brayer/vision

W. Orozco: Introduction to Digital Image Processing
www.geocities.com/CapeCanaveral/Campus/2165/